Multiparameter models

- Usually, we have models with many parameters, let's start with k=2.
  - **$p(\theta_1, \theta_2 | X) = p(X | \theta_1, \theta_2) p(\theta_1, \theta_2)$**
  - **$p(\theta_1, \theta_2)$ is joint prior. Often used: $p(\theta_1)\, p(\theta_2)$**
  - **Prior could also be hierarchical $p(\theta_1 | \theta_2)\, p(\theta_2)$**
  - **$p(X | \theta_1, \theta_2)$ could be e.g. $N(\mu,\sigma^2)$**

  - **Marginal posterior density**
  - **$p(\theta_1 | X) = \int p(\theta_1, \theta_2 | X)\, d\theta_2$**

    $= \int p(\theta_1 | \theta_2, X)\, p(\theta_2 | X)\, d\theta_2$

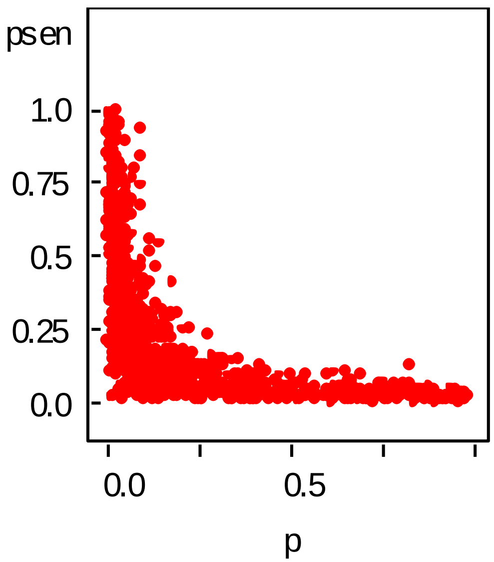
Multiparameter models

- The parameter of interest can be θ_1 while θ_2 is just a nuisance parameter.
  - **Example: diagnostic testing with sensitivity <100%**
  - **X ~ Bin(N, $\theta_1 * \theta_2$)**
  - Here, θ_1 is the unknown true prevalence, θ_2 is the unknown test sensitivity – for which we could have an informative prior, though.
  - We should take into account the uncertainty of both parameters jointly, given the data (and prior).
  - $p(\theta_1, \theta_2 \mid X) = \text{Bin}(X \mid N, \theta_1\theta_2)\, p(\theta_1)\, p(\theta_2)$

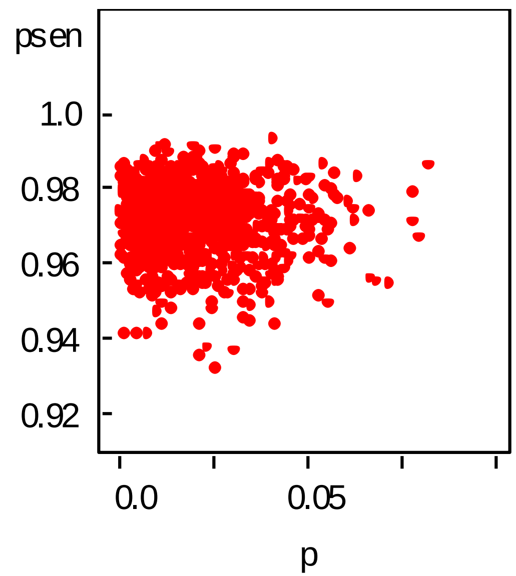
…Solving posterior is difficult, that's why WinBUGS is used…

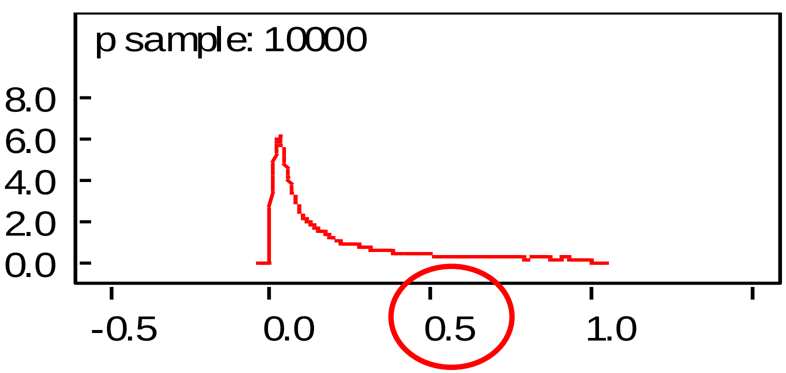
- **Assume we observed N=100, X=1.**

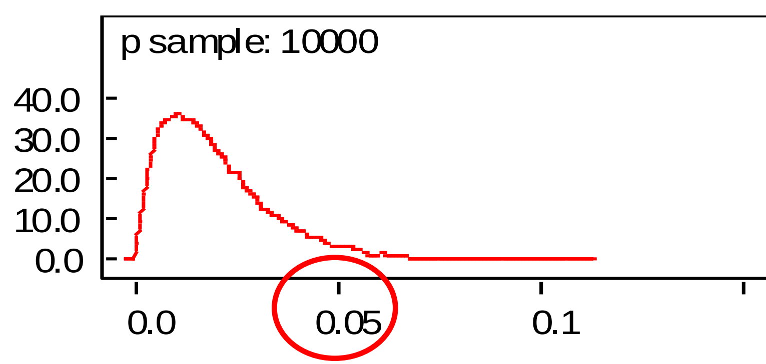
Without any prior knowledge of sensitivity →



Assuming sensitivity is average 0.97, SD 0.01 →



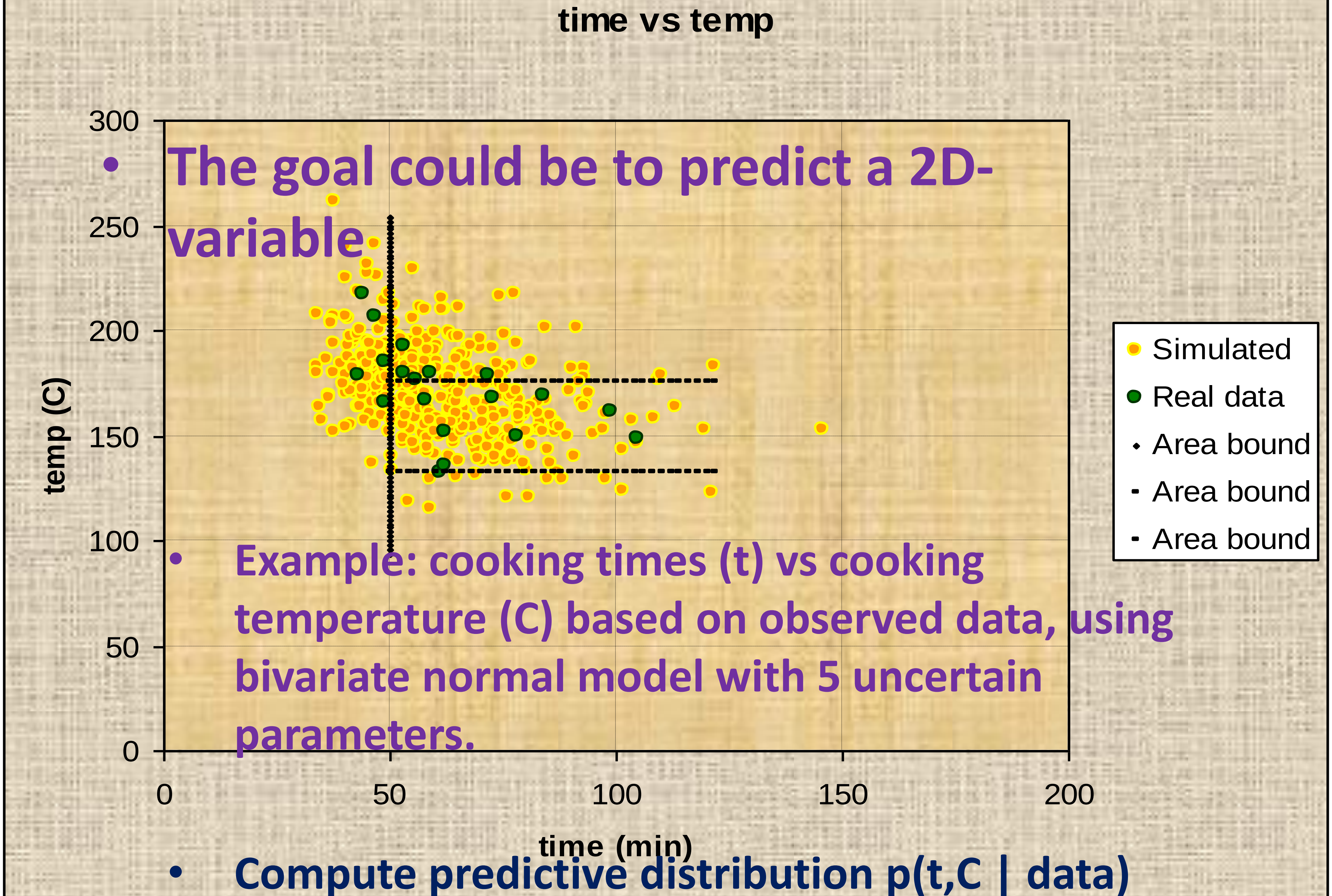




Multiparameter models

- The aim could also be to predict a multivariate response. (Correlated data models)
    - This requires several parameters in the model.
    - **$p(X_1, X_2 \mid \theta_1, \ldots, \theta_k)$**
    - Posterior prediction **$p(X_1^*, X_2^* \mid X_1, X_2)$** requires integration over all parameters
    - Then, some more integration to get marginal predictive distributions **$p(X_1^* \mid X_1, X_2) = \int p(X_1^*, X_2^* \mid X_1, X_2) dX_2^*$**

- The goal could be to predict a 2D-variable



- Example: cooking times (t) vs cooking temperature (C) based on observed data, using bivariate normal model with 5 uncertain parameters.

- Compute predictive distribution p(t,C | data)

Identifiability

- Parameters are unidentifiable (from data) if $P(X \mid \theta_1) = P(X \mid \theta_2)$, with $\theta_1 \neq \theta_2$
- Posterior result then depends solely on prior.
- Example: $X \sim N(\theta_1 + \theta_2, 1)$
  - All combinations with $\theta_1 + \theta_2 = c$ are equally probable, unless prior can make a difference.
  - Is the posterior a proper density?

  - Multiparameter models with insufficient data may lead to problems of identifiability. Useful to check the likelihood.

Multinomial model

- E.g. large bag of balls of k different colors. Pick N balls (with replacement)
- **$X_1,\ldots,X_k$ = number of balls of each color.**
- **$X_1+,\ldots,+X_k = N$**
- Vector X is multinomially distributed, given the true proportions $\theta_1,\ldots,\theta_k$.

- **Find out $p(\theta_1,\ldots,\theta_k \mid X)$**

Multinomial model

- This is a generalization of earlier inference problem with Binomial & Beta
- $p(\theta_1,\ldots,\theta_k) = \text{Dirichlet}(\alpha_1,\ldots,\alpha_k)$
- $\Sigma\theta_i = 1$
- Thanks to conjugate prior:

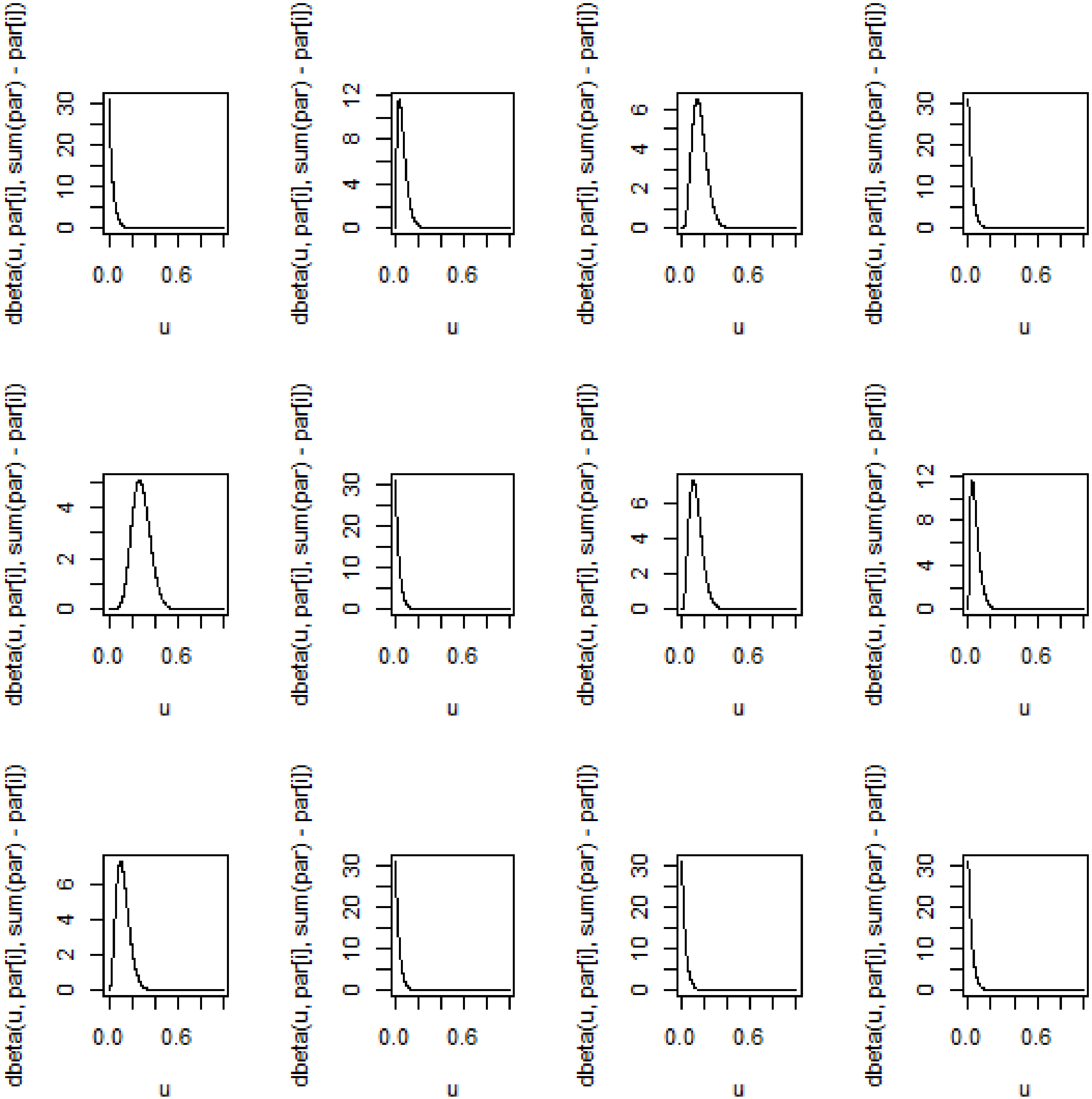
$$p(\theta_1,\ldots,\theta_k \mid X) = \text{Dirichlet}(\alpha_1+X_1,\ldots,\alpha_k+X_k)$$

- **Marginal densities easy, if $\theta \sim \text{Dir}(\alpha)$, then**

$$p(\theta_i \mid X) = \text{Beta}(\alpha_i, \Sigma\alpha_j - \alpha_i)$$

Multinomial model

- Example: there are 12 subtypes of bacteria. In a sample of 20, we observed the following numbers of each type:
- X=(0,1,4,0,8,0,3,1,3,0,0,0)
- **$p(\theta_1,\ldots,\theta_k \mid X) = Dir(\alpha_1+X_1,\ldots,\alpha_k+X_k)$**
- Note the ’pseudo data’ n=12 in the Dir(1,…,1) prior.


dbeta(u, par[i], sum(par) - par[i])
u
0.0
0.6
0
10
20
30
4
8
12
2
6

Normal model N(X|μ,σ)

- **Take a look at the easy cases first:**
- **p(μ|X,σ) and p(σ|X,μ)**
- Convenient notation: precision $\tau=1/\sigma^2$ this parameterization is also used in BUGS with normal densities.
- Conjugate prior for μ is $N(\mu_0,\sigma_0)$

$$p(\mu \mid \mu_0, \tau_0) = \exp(-0.5\tau_0(\mu-\mu_0)^2)/c$$

- Assume first a single observation X_i:

$$p(X_i \mid \mu, \tau) = \exp(-0.5\tau(X_i-\mu)^2)/c$$

Normal model N(X|**μ,σ**)

- **Posterior for μ is then**

$$p(\mu \mid X_i, \tau, \mu_0, \tau_0) = \exp(-0.5(\tau_0(\mu - \mu_0)^2 + \tau(X_i - \mu)^2))/c$$

$$= N\left(\frac{n_0\mu_0 + X_i}{n_0 + 1}, \frac{\sigma^2}{n_0 + 1} \right)$$

- Use ’completing a square’ –technique.
- Here $n_0 = \tau_0/\tau$ can be interpreted as ’pseudo sample size’ from the prior.
- **Posterior mean: $w\mu_0 + (1-w)X_i$, $w = \tau_0/(\tau_0 + \tau)$**

Normal model N(X|**μ,σ**)

- With several measurements $X_1,\dots,X_N$, we can write the data-model as

$$p(\bar{X} \mid \mu,\sigma) = N(\bar{X} \mid \mu,\sigma^2 / N)$$

- Similar to previous case, the posterior is

$$N\left(\frac{n_0\mu_0 + \bar{X}}{n_0+1}, \frac{\sigma^2 / N}{n_0+1}\right)$$

- Here $n_0 = \tau_0/(N\tau)$

Normal model N(X|μ,σ)

- Posterior mean and variance can also be expressed as

$$E(\mu \mid \overline{X}) = \frac{\dfrac{\mu_0}{\sigma_0^2} + \dfrac{N\overline{X}}{\sigma^2}}{\dfrac{1}{\sigma_0^2} + \dfrac{N}{\sigma^2}}$$

$$V(\mu \mid \overline{X}) = \frac{1}{\dfrac{1}{\sigma_0^2} + \dfrac{N}{\sigma^2}}$$

- What happens when N→0, or N→∞ ?

Normal model N(X|μ,σ)

- **Improper prior** $p(\mu) \propto 1$
- The posterior is proper density, and

$$p(\mu \mid \overline{X}) = N(\overline{X}, \sigma^2 / N)$$

- Compare with non-bayesian statistics, where the inference is based on

$$p(\overline{X} \mid \mu) = N(\mu, \sigma^2 / N)$$

- These are like mirror images…

Normal model N(X|**μ,σ**)

- **p(σ|X,μ) ?**
- Assume observations $X_1,\ldots,X_N$

$$p(X \mid \mu,\sigma) \propto \sigma^{-N} \exp(-\frac{1}{2\sigma^2}\sum_{i=1}^{N}(X_i-\mu)^2)$$

$$=(\sigma^2)^{-N/2}\exp(-\frac{N}{2\sigma^2}s_0^2) = \tau^{N/2}\exp(-\frac{N\tau}{2}s_0^2)$$

- Here $s_0^2 = \frac{1}{N}\sum_{i=1}^{N}(X_i-\mu)^2$
- Conjugate prior for τ? ....gamma(α,β)

Normal model N(X|μ,σ)

- Following from Bayes, the posterior p(σ|X,μ) is proportional to

$$\tau^{N/2}\exp(-\frac{N\tau}{2}s_0^2)\times\tau^{\alpha-1}\exp(-\beta\tau)$$

$$=\tau^{N/2+\alpha-1}\exp(-(\frac{N}{2}s_0^2+\beta)\tau)$$

- This is recognized as gamma($N/2+\alpha, Ns_0^2/2+\beta$)
- Uninformative prior $\alpha \to 0$, $\beta \to 0$.

Normal model $N(X|\mu,\sigma)$

- **$p(\mu,\sigma|X)$?**
- Assume observations $X_1,\ldots,X_N$
- Several options:

1. conjugate 2D prior $p(\mu,\sigma)=p(\mu|\sigma)p(\sigma)$
2. independent priors $p(\mu)$, $p(\sigma)$
3. improper prior $p(\mu,\tau)\propto 1/\tau$

This will get more mathematical, you are free to skip details unless you love the math…

Normal model $N(X|\boldsymbol{\mu},\boldsymbol{\sigma})$

- **Difficulties**:

1. conjugate 2D prior $p(\mu,\sigma)=p(\mu|\sigma)p(\sigma)$

Not very practical to express prior of μ, conditionally on σ.

Normal model N(X|μ,σ)

- **Difficulties**:

2. independent priors p(μ), p(σ)

Not possible to choose so that posterior could be solved in any familiar form.

Normal model N(X|$\boldsymbol{\mu},\boldsymbol{\sigma}$)

- **Difficulties**:

3. Improper prior $p(\mu,\tau) \propto 1/\tau$

same as $p(\mu,\sigma) \propto 1/\sigma^2$

same as $p(\mu,\log(\sigma)) \propto 1$

Posterior can be solved by factorization

$p(\mu,\sigma^2|X) = p(\mu|\sigma^2,X)p(\sigma^2|X)$

…we already have solved the first part before.

Normal model N(X|μ,σ)

- The second part is p(σ^2|X)

= Scaled-Inverse-χ^2(n-1,s)

$$s^2 = \frac{1}{N-1}\sum_{i=1}^{N}(X_i - \overline{X})^2$$

- Or: p(τ|X)

= Gamma((n-1)/2,(n-1)s^2/2)

- The full joint density can thus be written as a product of two known densities.
  - Convenient for Monte Carlo simulations. (draw σ^2, then μ conditionally on σ^2)
  - Also, can solve p(σ^2 |μ,X), useful for Gibbs sampling.

Working out $p(\sigma^2|X)$

- First, write $p(\mu,\sigma^2|\ X_1,\ldots,X_n\ )$ in the form:

prior **likelihood**

$$p(\mu,\sigma\,|\,X) \propto \sigma^{-2}\sigma^{-n}\exp\left(-\frac{1}{2\sigma^2}\sum_{i=1}^{N}(X_i-\mu)^2\right)$$

Manipulation

$$=\sigma^{-n-2}\exp\left(-\frac{1}{2\sigma^2}[(n-1)s^2+n(\bar{X}-\mu)^2]\right)$$

where $s^2=\frac{1}{N-1}\sum_{i=1}^{N}(X_i-\bar{X})^2$

- Then, integrate over μ to get marginal density.

Working out $p(\sigma^2|X)$

- Solving $p(\sigma^2|X)$: integrate the joint density $p(\sigma^2,\mu|X)$ over μ.

$$p(\sigma^2 \mid X) \propto \int_{-\infty}^{\infty} \sigma^{-n-2} \exp(-\frac{1}{2\sigma^2}[(n-1)s^2 + n(\bar{X}-\mu)^2])d\mu$$

$$= \sigma^{-n-2} \exp(-\frac{1}{2\sigma^2}(n-1)s^2) \times \int_{-\infty}^{\infty} \exp(-\frac{n}{2\sigma^2}(\bar{X}-\mu)^2)d\mu$$

$$= \sigma^{-n-2} \exp(-\frac{1}{2\sigma^2}(n-1)s^2) \times \sqrt{2\pi\sigma^2 / n}$$

$$\propto (\sigma^2)^{-(n+1)/2} \exp(-\frac{(n-1)s^2}{2\sigma^2})$$

= Scaled-Inverse-χ^2(n-1,s)

For $\tau=1/\sigma^2$: this is Gamma((n-1)/2,(n-1)s^2/2)

Working out $p(\sigma^2|X)$

- That required a few steps and manipulations…
- The lesson was:
  - To give you an impression of what kind of tricks and techniques are needed for exact solutions.
  - To see why and how the seemingly simple principle of Bayes theorem leads to increasingly complicated math which has been a major obstacle in practical Bayesian applications in the past.
  - To give motivation for the next sessions on Monte Carlo methods and WinBUGS/OpenBUGS.

Other multiparameter models

- Regression models, e.g. linear regression
  - **Example: $Y_i \sim N(\mu_i, \sigma^2)$**
  - **$\mu_i = \beta_1 X_{i1} + \ldots + \beta_k X_{ik} = X_i \beta$ (vector notation)**
  - **β = regression parameters.**
  - **X = matrix of explanatory variables.**
  - **Y = observations from i=1,…,n individuals.**
  - Aim to compute $p(\beta, \sigma^2 | Y, X)$ which is k+1 dimensional density.
  - Typical priors aim to be uninformative.
  - Posterior is then proper, if n>k, and the rank of X (number of linearly independent columns) is k. This is the case in most applications.

Other multiparameter models

- Regression models, e.g. linear regression
  - **Example: $Y_i \sim N(\mu_i,\sigma^2)$, assume σ^2 is 'known'.**
  - $p(\beta | Y,X,\sigma)$ can then be solved, and it is:
    $N((X^TX)^{-1} X^T Y , (X^TX)^{-1} \sigma^2)$
  - Here, posterior mean $(X^TX)^{-1} X^T Y$ is the same as max likelihood estimate (in this case it's also the least squares estimate) of β.

Multiparameter models

- Generalized linear models
  - **Example: $Y_i \sim Bin(N_i,\theta_i)$,**
  - Link function: **$logit(\theta_i) = \log(\theta_i/(1-\theta_i)) = X_i\,\beta$**
  - Prior **$p(\beta)$**
  - Posterior **$p(\beta|Y,X) = p(Y|X,\beta)p(\beta)/c =$**

    $\prod Bin(Y_i|N_i,\theta_i)p(\beta)/c$

Other multiparameter models

- Hierarchical models
  - **Example:**

    $Y_{ijk} \sim N(\mu_{ij},\sigma^2_{ij})$, result from patient k in hospital j, in district i.
  - **$\mu_{ij} \sim N(\phi_i,\sigma^2_i)$, mean of hospital j, in district i**
  - **$\phi_i \sim N(\theta,\sigma^2)$, mean of district i.**
  - **$\theta \sim N(0,10000)$ prior of 'grand mean'**
  - **Also need priors for variance components.**

  - Compute: **$p(\mu_{ij},\phi_i,\theta, \sigma^2_{ij}, \sigma^2_i, \sigma^2 | Y) =$**

$p(Y|\mu_i,,\sigma^2_{ij})\ p(\mu_{ij}|\phi_i, \sigma^2_i)\ p(\phi_i|\theta, \sigma^2)\ p(\theta)\ p(\sigma^2_{ij})\ p(\sigma^2_i)\ p(\sigma^2)\ /c$

Other multiparameter models

- Hierarchical models
  - Intuition: compare with a genetic model of a family tree: grand parents, parents, children. An observation from a child gives information about cousins too!
  - Also used in meta-analysis & evidence synthesis.
  - Also known as multilevel models.
  - Random effect models, mixed effects, spatial models, spatiotemporal models, applications are wide…

- **Hierarchical models… (more about DAGs later)**